Lighting Estimation and Adjustment for Facial Images

Xiaoyue Jiang, Xiaoyi Feng, Jun Wu, and Jinye Peng

Abstract For robust face detection and recognition, the problem of lighting variation is considered as one of the greatest challenges. Lighting estimation and adjustment is a useful way to remove the influence of illumination for images. Due to the different prior knowledge provided by a single image and image sequences, algorithms dealing with lighting problems are always different for these two conditions. In this chapter we will present a lighting estimation algorithm for a single facial image and a lighting adjustment algorithm for image sequences. To estimate the lighting condition of a single facial image, a statistical model is proposed to reconstruct the lighting subspace where only one image of each subject is required. For lighting adjustment of image sequences, an entropy-based optimization algorithm is proposed to minimize the difference between consequent images. The effectiveness of those proposed algorithms are illustrated on face recognition, detection and tracking tasks.

1 Introduction

Face detection, recognition and tracking are difficult due to variations caused by pose, expression, occlusion and lighting (or illumination), which make the distribution of face object highly nonlinear. Lighting is regarded as one of the most critical factors for robust face recognition. Current attempt to handle lighting variation is either to find the invariant features and representations or to model the variation. The gradient based algorithm [1, 2] attempted to find illumination invariant features. While the other kind of algorithms tried to extract the reflectance

X. Jiang (✉)
School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China
e-mail: xjiang@nwpu.edu.cn

X. Feng • J. Wu
Northwestern Polytechnical University, Xi'an 710072, China

J. Peng
Northwestern University, Xi'an 710069, China

M. Kawulok et al. (eds.), *Advances in Face Detection and Facial Image Analysis*,
DOI 10.1007/978-3-319-25958-1_3

information from the observations in different lighting conditions, including the algorithms based on Quotient Images [3–8] and the algorithms based on Retinex theory [9–12]. But these methods cannot extract sufficient features for accurate recognition.

Early work on modeling lighting variation [13, 14] showed that a 3D linear subspace can represent the variation of a Lambertian object under a fixed pose when there is no shadow. With the same Lambertian assumption, Belhumeur and Kriegman [15] showed that images illuminated by an arbitrary number of point light sources formed a convex polyhedral cone, i.e. the illumination cone. In theory, the dimensionality of the cone is finite. They also pointed out that the illumination cone can be approximated by a few properly chosen images. Good recognition results of the illumination cone in [16] demonstrated its representation for lighting variation.

Recent research is mainly focused on the application of low-dimensional subspace to lighting variation modeling. With the assumption of Lambertian surface and non-concavity, Zhou et al. [17] found a set of basis images from training images based on photometric methods, while Ramamoorith and Hanrahan [18] and Basri and Jacobs [19] independently introduced the spherical harmonic (SH) subspace to approximate the illumination cone. Chen et al. [20] decomposed the original image into lighting and reflectance maps and build up lighting subspace from lighting maps. Sparse representation-based algorithm also showed its effectiveness in dealing with illumination problem [21–24], while the lack of training images always limited its application. In order to construct the lighting subspace, a lot of algorithms also applied the 3D model of faces to handling lighting variations [25–30]. However, recovering the 3D information from images is still an open problem in computer vision.

Lee et al. [31] built up a subspace that is nearest to the SH subspace and has the largest intersection with the illumination cone, called the nine points of light (9PL) subspace. It has a universal configuration for different subjects, i.e. the subspace is spanned by images under the same lighting conditions for different subjects. In addition, the basis images of 9PL subspace can be duplicated in real environments, while those of the SH subspace cannot because its the basis images contain negative values. Therefore the 9PL subspace can overcome the inherent limitation of SH subspace. Since the human face is neither completely Lambertain nor entirely convex, SH subspace can hardly represent the specularities or cast shadows (not to mention inter-reflection). The basis images of 9PL subspace are taken from real environment, they already contain all the complicated reflections of the objects. Therefore the 9PL subspace can give a more detailed and accurate description of lighting variation.

In practice, the requirement of these nine real images cannot always be fulfilled. Usually there are fewer gallery images (e.g. one gallery image) per subject, which can be taken under arbitrary lighting conditions. In this chapter, we propose a statistical model for recovering the 9 basis images of the 9PL subspace from only one gallery image. Based on the estimation of lighting coefficient for a facial image, its basis images can be constructed with the prior knowledge about the distribution of the basis images. An overview of the proposed basis construction algorithm

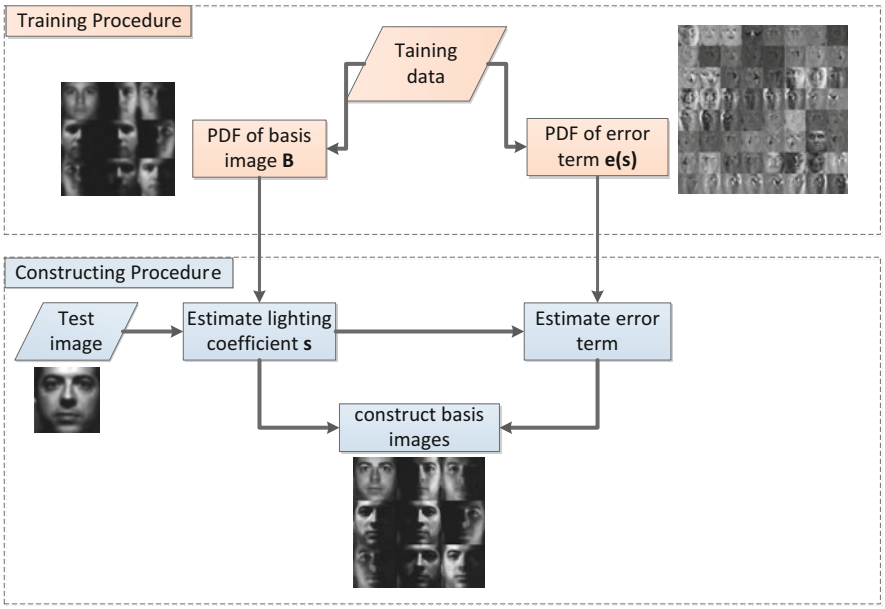

Fig. 1 Flowchart of constructing basis images for lighting subspace

is shown in Fig. 1. Zhang and Samaras [25] presented a statistical method for recovering the basis images of SH subspace instead. In their training procedure, geometric and albedo information is still required for synthesizing the harmonic images. In contrast, the proposed method requires only some real images that can be easily obtained in real environment. Since the recovered basis images of the 9PL subspace contain all the reflections caused by the shape of faces, such as cast shadows, specularities, and inter-reflections, better recognition results are obtained, even under extreme lighting conditions. Compared with other algorithms based on 3D model [25, 28, 29], the proposed algorithm is entirely a 2D algorithm, which has much lower computational complexity. For more details of the algorithm is presented in paper [32].

Lighting variations will also bring challenges for the processing of image sequences. Even though there are a lot of algorithms to deal with lighting problem for images, there are a few investigations for image sequences particularly. In fact most lighting algorithms are too complicated for image sequences, and do not consider the special characteristic of sequences. Also, the scene can be more complex in real applications. Thus it is impossible to build lighting models for different kind of subjects and adjust the lighting condition for each of them independently. However, the lighting adjustment can be applied to the whole scene according to an optimal lighting condition. In this chapter, we also propose a two-step entropy-based lighting adjustment algorithm for image sequences. An overview of the adjustment algorithm is shown in Fig. 2. Paper [33] presents more details about the lighting adjustment algorithm.

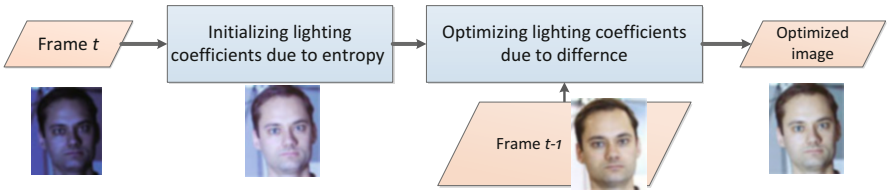

Fig. 2 Flowchart of lighting adjustment algorithm for image sequences

In the lighting adjustment algorithm, the difference between successive frames is used as a cost function for optimizing a lighting model of an image sequence. For each image frame, to ensure convergence, the proposed lighting adjustment algorithm works in two steps. First initial values of the lighting model parameters are estimated using the entropy of the current image as measure. These values are then used as initial guesses for a constrained least squares optimization problem, considering two successive frames. It is worth pointing out that the proposed algorithm did not only try to do shadow removal, which is only one aspect of lighting problem because there is no shadow but highlight or un-uniform lighting distribution in an image sometimes. Also, we adjusted the global lighting conditions to be more uniform and enhance the local features of the image as well.

This chapter is organized as follows. In Sect. 2, we briefly summarize the methods of low-dimensional linear approximation of the illumination cone, including the SH subspace and the 9PL subspace. The training of our statistical model and the application of the model for recovering basis images from only one gallery image are described in Sects. 3 and 4 respectively. Section 5 is dedicated to the experimental results. Then perception based lighting model is introduced in Sect. 6. In Sect. 7 we give an overview of the proposed lighting adjustment algorithm for image sequence. Section 8 discusses qualitative and quantitative results of the lighting adjustment in the case of facial features detection and tracking. Finally, conclusions are drawn in Sect. 9.

2 Approximation of the Illumination Cone

Belhumeur and Kriegman [15] proved that the set of n-pixel images of a convex object that had a Lambertian surface illuminated by an arbitrary number of point light sources at infinity formed a convex polyhedral cone, called the illumination cone $\mathscr{C}$ in $\mathscr{R}^n$. Each point in the cone is an image of the object under a particular lighting condition, and the entire cone is the set of images of the object under all possible lighting conditions. Any images in the illumination cone $\mathscr{C}$ (including the boundary) can be determined as a convex combination of extreme rays (images) given by

$$I_{ij} = max(\tilde{B}\tilde{s}_{ij}, 0) \tag{1}$$

where $\tilde{s}_{ij} = \tilde{b}_i \times \tilde{b}_j$ and $\tilde{B} \in \Re^{n\times 3}$. Every row of $\tilde{B}$, $\tilde{b}_i$, is a three element row vector determined by the product of the albedo with the inward pointing unit normal vector of a point on the surface. There are at most $q(q-1)$ extreme rays for $q \leq n$ distinct surface normal vectors. Therefore the cone can be constructed with finite extreme rays and the dimensionality of the lighting subspace is finite. However, building the full illumination cone is tedious, and the low dimensional approximation of the illumination cone is applied in practice.

From the view of signal processing, the reflection equation can be considered as the rotational convolution of incident lighting with the albedo of the surface [18]. The spherical harmonic functions $Y_{lm}(\theta, \phi)$ are a set of orthogonal basis functions defined in the unit sphere, given as follows,

$$Y_{lm}(\theta, \phi) = N_{lm} P_l^m(\cos\theta) \exp^{im\phi} \tag{2}$$

where $N_{lm} = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}$, (θ, ϕ) is the spherical coordinate (θ is the elevation angle, which is the angle between the polar axis and the z-axis with range $0 \leq \theta \leq 180°$, and ϕ is the azimuth angle with the range $-180° \leq \phi \leq 180°$). P_l^m is the associated Legendre function, and the two indices meet the conditions $l \geq 0$ and $l \geq m \geq -l$. Then functions in the sphere, such as the reflection equation, can be expanded by the spherical harmonic functions, which are basis functions on the sphere. Images can be represented as a linear combination of spherical harmonic functions. The first three order ($l \leq 3$) basis can account for 99 % energy of the function. Therefore the first three order basis functions (altogether 9) can span a subspace for representing the variability of lighting. This subspace is called the spherical harmonic (SH) subspace.

Good recognition results reported in [19] indicates that the SH subspace $\mathscr{H}$ is a good approximation to the illumination cone $\mathscr{C}$. Given the geometric information of a face, its spherical harmonic functions can be calculated with Eq. (2). These spherical harmonic functions are synthesized images, also called harmonic images. Except the first harmonic image, all the others have negative values, which cannot be obtained in reality. To avoid the requirement of geometric information, Lee et al. [31] found a set of real images which can also serve as a low dimensional approximation to illumination cone based on linear algebra theory.

Since the SH subspace $\mathscr{H}$ is good for face recognition, it is reasonable to assume that a subspace $\mathscr{R}$ close to $\mathscr{H}$ would be likewise good for recognition. $\mathscr{R}$ should also intersect with the illumination cone $\mathscr{C}$ as much as possible. Hence a linear subspace $\mathscr{R}$ which is meant to provide a basis for good face recognition will also be a low dimensional linear approximation to the illumination cone $\mathscr{C}$. Thus subspace should satisfy the following two conditions [31]:

1. The distance between $\mathscr{R}$ and $\mathscr{H}$ should be minimized.
2. The unit volume ($vol(\mathscr{C} \cap \mathscr{R})$) of $\mathscr{C} \cap \mathscr{R}$ should be maximized (the unit volume is defined as the volume of the intersection of $\mathscr{C} \cap \mathscr{R}$ with the unit ball).

Note that $\mathscr{C} \cap \mathscr{R}$ is always a subcone of $\mathscr{C}$; therefore maximizing its unit volume is equivalent to maximize the solid angle subtended by the subcone $\mathscr{C} \cap \mathscr{R}$. If $\{\tilde{I}_1, \tilde{I}_2, \cdots, \tilde{I}_k\}$are the basis images of $\mathscr{R}$. The cone $\mathscr{R}_c \subset \mathscr{R}$ is defined by $\tilde{I}_k$,

$$\mathscr{R}_c = \{I | I \in \mathscr{R}, I = \sum_{k=1}^{M} \alpha_k \tilde{I}_k, \alpha_k \geq 0\} \tag{3}$$

is always a subset of $\mathscr{C} \cap \mathscr{R}$. In practice the subcone $\mathscr{C} \cap \mathscr{R}$ is taken as $\mathscr{R}_c$ and the subtended angle of $\mathscr{R}_c$ is maximized. $\mathscr{R}$ is computed as a sequence of nested linear subspace $\mathscr{R}_0 \subseteq \mathscr{R}_1 \subseteq \cdots \subseteq \mathscr{R}_i \subseteq \cdots \subseteq \mathscr{R}_9 = \mathscr{R}$, with $\mathscr{R}_k (k > 0)$ a linear subspace of dimension i and $\mathscr{R}_0 = \emptyset$. First, EC denotes the set of (normalized) extreme rays in the illumination cone $\mathscr{C}$; and EC_k denotes the set obtained by deleting k extreme rays from EC, where $EC_0 = EC$. With $\mathscr{R}_{k-1}$ and EC_{k-1}, the sets EC_k and R_k can be defined iteratively as follows:

$$\tilde{I}_k = \arg \max_{I \in EC_{k-1}} \frac{dist(I, \mathscr{R}_{k-1})}{dist(I, \mathscr{H})} \tag{4}$$

where $\tilde{I}_k$ denotes the element in EC_{k-1}. $\mathscr{R}_k$ is defined as the space spanned by $\mathscr{R}_{k-1}$ and $\tilde{I}_k$. $EC_k = EC_{k-1} \backslash \tilde{I}_k$. The algorithm stops when $\mathscr{R}_9 \equiv \mathscr{R}$ is reached. The result of Eq. (4) is a set of nine extreme rays that span $\mathscr{R}$ and there are nine directions corresponding to these nine extreme rays. For different subjects, the nine lighting directions are qualitatively very similar. By averaging Eq. (4) of different subjects and maximizing this function as follows:

$$\tilde{I}_k = \arg \max_{I \in EC_{k-1}} \sum_{p=1}^{N} \frac{dist(I^p, \mathscr{R}^p_{k-1})}{dist(I^p, \mathscr{H}^p)} \tag{5}$$

where I^p denotes the image of subject p taken under a single light source. H^p is the SH subspace of subject p. $\mathscr{R}^p_{k-1}$ denotes the linear subspace spanned by images $\{\tilde{I}^p_1, \cdots, \tilde{I}^p_k\}$ of subject p. The universal configuration of nine light source direction is obtained. They are $(0, 0)$, $(68, -90)$, $(74, 108)$, $(80, 52)$, $(85, -42)$, $(85, -137)$, $(85, 146)$, $(85, -4)$, $(51, 67)$. The directions are expressed in spherical coordinates as pairs of (ϕ, θ), Fig. 3a illustrates the nine basis images of a person from the Yale face database B [16].

3 Statistical Model of Basis Images

According to the universal configuration of lighting directions, we can apply nine images taken under controlled environment to spanning the 9PL linear subspace. However, even these nine images may not be available in some situations. Thus, we propose a statistical method for estimating the basis images from one gallery image.

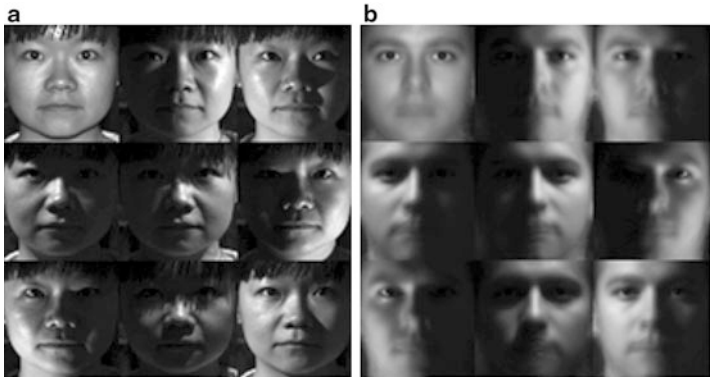

Fig. 3 The basis images of 9PL subspace. (**a**) images taken under certain lighting conditions can serve as the basis images of the object. (**b**) The mean images of the basis images estimated from the bootstrap data set

To build the statistical model, we must find the probability density function (PDF) of basis images and the PDF of the error term. Due to the limited amount of the training data, we use the bootstrap method to estimate the statistics of basis images. The recovering step is to estimate the corresponding basis images from one single image of a novel subject under arbitrary lighting conditions. For a given image, we first estimate its lighting coefficient. Then according to the maximum a posteriori (MAP) estimation, we obtain an estimation of the basis images. Finally, we apply the recovered subspace to face recognition. The probe image is identified as the face whose lighting subspace is closest in distance to the image.

Given nine basis images, we can construct images under arbitrary lighting conditions as follows,

$$I = B\mathbf{s} + e(\mathbf{s}) \tag{6}$$

where $I \subset \Re^{d\times 1}$ is the image vector. $B \subset \Re^{d\times 9}$ is the matrix of nine basis images, every column of which is the vector of the basis image. $\mathbf{s} \subset \Re^{d\times 1}$ is the vector of lighting coefficients which denotes the lighting conditions of the image. Error term $e(\mathbf{s}) \subset \Re^{d\times 1}$ is related to the pixels' position and lighting conditions.

For a novel image, we estimate its basis images through the maximum a posterior (MAP) estimation. That is

$$B_{MAP} = \arg\max_{B} P(B|I) \tag{7}$$

According to the Bayes rule

$$P(B|I) = \frac{P(I|B)P(B)}{P(I)} \tag{8}$$

where $P(I)$ is the evidence factor which guarantees that posterior probabilities would sum to one. Then Eq. (7) can become

$$B_{MAP} = \arg\max_{B}(P(I|B)P(B)) \tag{9}$$

In order to recover the set of basis images from an image with Eq. (9), one should know the PDF of the basis images, i.e. $P(B)$, and the PDF of the likelihood, i.e. $P(I|B)$. Assuming the error term of Eq. (6) is normally distributed with mean $\mu_e(x, \mathbf{s})$ and variance $\sigma_e^2(x, \mathbf{s})$, we can deduce that the PDF of the likelihood $P(I|B)$, is also Gaussian with mean $B\mathbf{s} + \mu_e(\mathbf{s})$ and variance $\sigma_e^2(\mathbf{s})$ according to Eq. (6).

We assume that the PDF of the basis images B are Gaussians of unknown means μ_B and covariances C_B as in [25, 34]. The probability $P(B)$ can be estimated from the basis images of the training set. In our experiments, the basis images of 20 different subjects from the extended Yale face database B [16] are introduced to the bootstrap set. Note that, the basis images of every subject are real images which were taken under certain lighting conditions. The lighting conditions are determined by the universal configurations of the 9PL subspace. The sample mean μ_B and sample covariance matrix $C_B$are computed. Figure 3b shows the mean basis images, i.e. μ_B.

The error term $e(\mathbf{s}) = I - B\mathbf{s}$ models the divergence between the real image and the estimated image which is reconstructed by the low dimensional subspace. The error term is related to the lighting coefficients. Hence, we need to know the lighting coefficients of the training images before estimating the error term. For a training image, its lighting coefficient can be estimated by solving the linear equation $I = B\mathbf{s}$. For every subject in the extended Yale face database B, there are 64 images under different lighting conditions. We use the images of the same 20 subjects whose basis images are used for training the mean value of basis, i.e. μ_B, for computing the statistics of the error. Under a certain lighting condition, we estimate the lighting coefficients of every subject's image, i.e. $\mathbf{s}_k^p$ (the lighting coefficients of the p^{th} subject's image under the lighting condition $\mathbf{s}_k$. The mean value of different subjects' lighting coefficients can be the estimated coefficients ($\bar{\mathbf{s}}_k$) for that lighting condition, i.e. $\bar{\mathbf{s}}_k = \sum_{p=1}^{N} \mathbf{s}_k^p / N$. Then, under a certain lighting condition, the error term the of the p^{th} subject's image is

$$e_p(\bar{\mathbf{s}}_k) = I_k^p - B_p\bar{\mathbf{s}}_k \tag{10}$$

where I_k^p is the training image of the p^{th} subject under lighting condition $\mathbf{s}_k$ and B_p is the basis images of the p^{th} subject. Following the above assumption, we can estimate the mean $\mu_e(\bar{\mathbf{s}}_k)$ and variance $\sigma_e^2(\bar{\mathbf{s}}_k)$ of the error term.

4 Estimating the Basis Images

As described in the previous section, the basis images of a novel image can be recovered by using the MAP estimation. But before applying the statistical model to estimating the basis images, one needs to first estimate the lighting coefficient **s** of the image. Then, the error term under those lighting conditions can be estimated.

4.1 Estimating Lighting Coefficients

Lighting influences greatly the appearance of an image. Under similar illumination, images of different subjects will appear almost the same. The difference between the images of the same subject under different illuminations is always larger than that between the images of different subjects under the same illumination [35]. Therefore we can estimate the lighting coefficients of a novel image with an interpolation method. The kernel regression is a smooth interpolation method [36]. It is applied to estimating the lighting coefficients. For every training image, we have their corresponding lighting coefficients. For a novel image I_n, its lighting coefficient is given by

$$\mathbf{s} = \frac{\sum_{k=1}^{M} w_k \mathbf{s}_k^p}{\sum_{k=1}^{M} w_k} \tag{11}$$

$$w_k = \exp(-\frac{[D(I_n, I_k^p)]^2}{2(\sigma_{I_k^p})^2}) \tag{12}$$

where $D(I_n, I_k^p) = \|I_n - I_k^p\|_2$ is the L_2 norm of the image distance. $\sigma_{I_k^p}$ determines the weight of test image I_k^p in the interpolation. M is the number of images that have similar lighting condition. In the training set, every subject has 64 different images and there are altogether 20 different subjects. Thus, for a novel image, there will be $M = 20$ images with similar illumination. In our experiment, we assign the farthest distance of these 20 images from the probe image to $\sigma_{I_k^p}$. $\mathbf{s}_k^p$ is the lighting coefficient of image I_k^p.

4.2 Estimating the Error Term

The error term denotes the difference between the reconstructed image and the real image. This divergence is caused by the fact that the 9PL subspace is the low-dimensional approximation to the lighting subspace, and it only accounts for the

low frequency parts of the lighting variance. Given the statistics of the error term under known illumination, i.e. $\mu_e(\bar{\mathbf{s}}_k)$, $\sigma_e^2(\bar{\mathbf{s}}_k)$, those under a new lighting condition can be estimated, also via the kernel regression method [34].

$$\mu_e(\mathbf{s}) = \frac{\sum_{k=1}^{M} w_k \mu_e(\bar{\mathbf{s}}_k)}{\sum_{k=1}^{M} w_k} \tag{13}$$

$$\sigma_e^2(\mathbf{s}) = \frac{\sum_{k=1}^{M} w_k \sigma_e^2(\bar{\mathbf{s}}_k)}{\sum_{k=1}^{M} w_k} \tag{14}$$

$$w_k = \exp(-\frac{[D(\mathbf{s}, \bar{\mathbf{s}}_k)]^2}{2[\sigma_{\bar{\mathbf{s}}_k}]^2}) \tag{15}$$

where $D(\mathbf{s}, \bar{\mathbf{s}}_k) = \|\mathbf{s} - \bar{\mathbf{s}}_2\|_2$ is the L_2 norm of the lighting coefficient distance. Like $\sigma_{I_k^p}$, $\sigma_{\bar{\mathbf{s}}_k}$ determines the weight of the error term related to the lighting coefficients $\bar{\mathbf{s}}_k$. Also, we assign the farthest lighting coefficient distance of these 20 images from the probe image to $\sigma_{\bar{\mathbf{s}}_k}$.

4.3 Recovering the Basis Images

Given the estimated lighting coefficients $\mathbf{s}$ and the corresponding error term $\mu_e(\mathbf{s})$, $\sigma_e^2(\mathbf{s})$, we can recover the basis images via the MAP estimation. If we apply the log probability, omit the constant term, and drop $\mathbf{s}$ for compactness, Eq. (9) can become

$$\arg\max_B \left(-\frac{1}{2}(\frac{I - B\mathbf{s} - \mu_e}{\sigma_e})^2 - \frac{1}{2}(B - \mu_B)C_B^{-1}(B - \mu_B)^T \right) \tag{16}$$

To solve Eq. (16), we estimate the derivatives,

$$-\frac{2}{\sigma_e^2}(I - B\mathbf{s} - \mu_e)\mathbf{s}^T + 2(B - \mu_B)C_B^{-1} = 0 \tag{17}$$

Then we rewrite Eq. (17) as a linear equation,

$$AB = b \tag{18}$$

where $A = \frac{\mathbf{s}\mathbf{s}^T}{\sigma_e^2} + C_B^{-1}$ and $b = \frac{I - \mu_e}{\sigma_e^2}\mathbf{s} + C_B^{-1}\mu_B$.The solution of the linear equation is $B = A^{-1}b$. Using the Woodbury's identity [37], we can obtain an explicit solution

$$B_{MAP} = A^{-1}b$$

$$= \left(C_B - \frac{C_B \mathbf{s}\mathbf{s}^T C_B}{\sigma_e^2 + \mathbf{s}^T C_B \mathbf{s}} \right) \left(\frac{I - \mu_e}{\sigma_e^2} \mathbf{s} + C_B^{-1} \mu_B \right)$$

$$= \left(\frac{I - \mu_B \mathbf{s} - \mu_e}{\sigma_e^2 + \mathbf{s}^T C_B \mathbf{s}} \right) C_B \mathbf{s} + \mu_B \tag{19}$$

From Eq. (19), the estimated basis image is composed of the term of characteristics, $\left(\frac{I - \mu_B \mathbf{s} - \mu_e}{\sigma_e^2 + \mathbf{s}^T C_B \mathbf{s}} C_B \mathbf{s} \right)$, and the term of mean, μ_B. In the term of characteristics, $(I - \mu_B \mathbf{s} - \mu_e)$ is the difference between the probe image and the images reconstructed via the mean basis images.

4.4 Recognition

The most direct way to perform recognition is to measure the distance between probe images and the subspace spanned by the recovered basis images. Every column of B is one basis image. However, the basis images are not orthonormal vectors. Thus we perform the QR decomposition on B to obtain a set of orthonormal basis, i.e. the matrix Q. Then the projection of probe image I to the subspace spanned by B is $QQ^T I$, and the distance between the probe image I and the subspace spanned by B can be computed as $\|QQ^T I - I\|_2$. In the recognition procedure, the probe image is identified as the subspace with minimum distance from it.

5 Experiments on Lighting Estimation

The statistical model is trained by images from the extended Yale face database B [16]. With the trained statistical model, we can reconstruct the lighting subspace from only one gallery image, which should be insensitive to lighting variation. Thus, recognition can be achieved across illumination conditions.

5.1 Recovered Basis Images

To recover the basis images from a single image, the lighting coefficients of the image should be estimated first. Then we estimate the error terms of the image. Finally, the basis images of the image can be obtained with Eq. (19).

Although the images of the same object are under different lighting conditions, the recovered basis images should be similar. The probe images are from the Yale face database B. There are 10 subjects and 45 probe images per subject. According to the lighting conditions of the probe images, they can be grouped into 4 subsets as

Table 1 The subsets of Yale face database B

	subset1	subset2	subset3	subset4
Illumination	0–12	13–25	26–50	50–77
Number of images	70	120	120	140

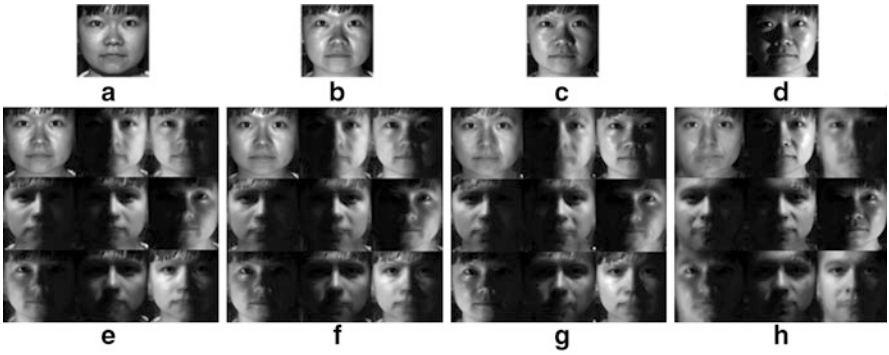

Fig. 4 Recovered basis images. (**a**)–(**d**) are images in subset 1–4 of Yale face database B respectively. (**e**)–(**h**) are recovered basis images from image (**a**)–(**d**) respectively

in [16]. The details can be found in Table 1. From subset1 to subset4, the lighting conditions become extreme. For every subject, we recover its basis images from only one of its probe images each time. Then we can obtain 45 sets of basis images for every subject. Figure 4 shows the basis images recovered from an image of each subset. $\bar{\sigma}_{basis}$ (the mean standard deviation of the 45 sets of basis images of 10 subjects) is 7.76 intensity levels per pixel, while $\bar{\sigma}_{image}$ (the mean standard deviation of the original 45 probe images of 10 subjects) is 44.12 intensity levels per pixel. From the results, we can see that the recovered basis images are insensitive to the variability of lighting. Thus we can recover the basis images of a subject from its images under arbitrary lighting conditions.

5.2 Recognition

Recognition is performed on the Yale face database B [16] first. We take the frontal images (pose 0) as the probe set, which is composed of 450 images (10 subjects, 45 images per subject). For every subject, one image is used for recovering its lighting subspace and the 44 remaining images are used for recognition. The comparison of our algorithm with the reported results is shown in Table 2.

Our algorithm reconstructed the 9PL subspace for every subject. The recovered basis images also contained complicated reflections on faces, such as cast shadows, specularities, and inter-reflection. Therefore the recovered 9PL subspace can give a more detailed and accurate description for images under different lighting conditions. As a result, we can get good recognition results on images with different

Table 2 The recognition error rate of different recognition algorithms on Yale face database B

Algorithms	subset1 & 2	subset3	subset4
Correlation [16]	0.0	23.3	73.6
Eigenfaces [16]	0.0	25.8	75.7
LTV [9]	0.2	21.4	24.5
S&L(LOG-DCT) [5]	0.0	14.0	15.7
S&L(NPL-QI) [5]	0.0	3.2	13
Linear subspace [16]	0.0	0.0	15.0
Cones-attached [16]	0.0	0.0	8.6
Cones-cast [16]	0.0	0.0	0.0
harmonic images-cast [16]	0.0	0.0	2.7
3D based SH model [25]	0.0	0.3	3.1
SH model extreme [25]	7.75	7.5	6.8
BIM (30 bases) [28]	0.0	0.0	0.7
Wang et al. [29]	0.0	0.0	0.1
9PL-real [31]	0.0	0.0	0.0
Intrinsic lighting subspace [20]	0.0	0.0	5.71
Our algorithm	0.0	0.0	0.72

lighting conditions. Also, the reported results of 'cone-cast', 'harmonic images-cast' and '9PL-real' showed that better results can be obtained when cast shadows were considered. Although paper [28, 29] also use one image to adjust lighting conditions, they need to recover the 3D model of the face first. Our algorithm is a completely 2D-based approach. Computationally, it is much less expensive compared with those 3D based methods. The basis images of a subject can be directly computed with Eq. (19) while the recognition results are comparable to those from the 3D-based methods. Compared with the results of Quotient-based methods [5, 9], the results of lighting-subspace-based methods [20, 25, 28, 29, 31] are much better. That is due to the loss of appearance information in Quotient-based methods. Actually, these information contained a lot of low-frequency information which is important for recognition.

5.3 Multiple Lighting Sources

An image taken under multiple lighting sources can be considered as images taken under a single lighting source being superimposed. Through interpolation, the lighting coefficients of images taken under single lighting are linearly combined to approximate those of the image taken under multiple-lighting. Here we also apply the statistical model trained on the extended Yale Database B to basis images estimation. Similarly the lighting coefficients of images are estimated through

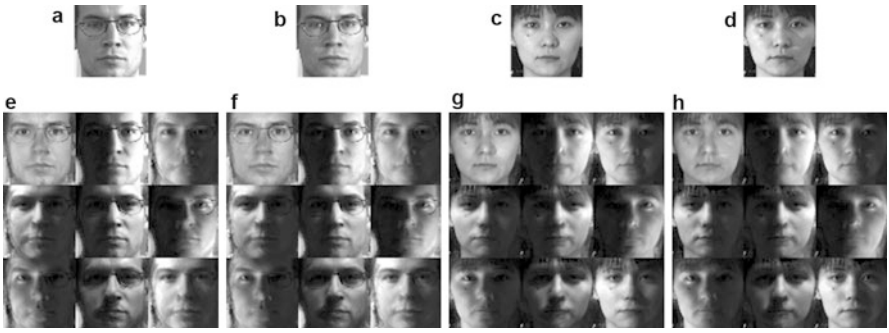

Fig. 5 Recovered basis images. (**a**) and (**b**) are images in PIE database, (**e**) and (**f**) are estimated basis images from image (**a**) and (**b**), respectively. (**c**) and (**d**) are images in AR database, (**g**) and (**h**) are estimated basis images from image (**c**) and (**d**), respectively

Table 3 Recognition rate on different databases

Face database	PIE	AR
$\bar{\sigma}_{basis}$	11.01	11.34
$\bar{\sigma}_{image}$	285	38.59
Recognition rate (%)	98.21	97.75

interpolation. Then the error term can be estimated according to the lighting coefficients. Finally, the basis images are recovered.

In the PIE face database [38], there are 23 images per subject taken under multiple lighting sources, and altogether 69 subjects. We recover 23 sets of the basis images from the 23 images of every subject respectively. With these estimated basis images, we perform recognition on the 1587 images (23 images per person) 23 times. We also estimate basis for images in the AR database [39]. We select randomly 4 images under different illumination per subject (image 1, 5, 6, 7) and recover the respective basis images from those images. Recognition is performed on 504 images (126 subjects and 4 images per subject) 4 times. Samples of the recovered basis images from images in the PIE and ARdatabases are shown in Fig. 5. The average recognition rates, the mean standard deviation of the recovered basis images ($\bar{\sigma}_{basis}$) and the mean standard deviation of the gallery images ($\bar{\sigma}_{images}$) are presented in Table 3. The results show that the statistical model trained by images taken under single lighting source can also be generalized to images taken under multiple lighting sources. The recognition results on PIE dataset is also compared with some lighting pre-processing methods reported in paper [40], as shown in Table 4. From the results, we can see that the proposed algorithm performed better than those reflectance estimation methods [9–12]. That is due to the accurate description of appearance in different lighting conditions with estimated basis images.

Table 4 Comparison with other lighting preprocessing methods on the dataset of PIE

Algorithm	LTV [9]	TT [10]	LDCT [11]	LN [12]	Our algorithm
Recognition rate	89.1	94.4	92.1	97.1	98.21

6 Perception-Based Lighting Model

The Human Vision System (HVS) can adapt very well under enormously changed lighting conditions. People can see well at daytime and also at night. That is due to the accurate adaptation ability of the HVS. However, image capturing devices seldom have this adaptation ability. For an image taken under extreme lighting conditions, such as the images shown in first row of Fig. 7b, a proper lighting adjustment algorithm should not only adjust the brightness of the images, but also enhance the features of the image, especially for the dark regions. To reach this goal, we propose to reduce the light variations by an adaptive adjustment of the image. Here, we employ a model of photoreceptor adaptation in Human Vision System [41] in which three parameters (α, f, m) control the lighting adjustment. The adjusted image Y is modeled as a function of these lighting parameters and the input image X as:

$$Y(\alpha, m, f; X) = \frac{X}{X + \sigma(X_a)} V_{max} \tag{20}$$

where σ, referred to as *semi-saturation constant*, X_a *the adaptation level*, and V_{max} determines the maximum range of the output value (we use $V_{max} = 255$ to have grey image output in the range of $[0, 255]$). The semi-saturation constant σ describes the image intensity and its contrast through the parameters f and m, respectively [41]:

$$\sigma(X_a) = (fX_a)^m \tag{21}$$

Adaptation Level I_a If we choose the average intensity of the image as the adaptation level I_a, the adjustment is global. It does not perform any specific processing to the darker or brighter region and some details in those regions may be lost. To compensate the details, the local conditions of every point should be considered. We can use the bi-linear interpolation to combine the global adaptation I_a^{global} and local adaptation $I_a^{local}(x, y)$ as,

$$I_a(x, y) = \alpha I_a^{local}(x, y) + (1 - \alpha) I_a^{global} \tag{22}$$

$$I_a^{local}(x, y) = K(I(x, y)) \tag{23}$$

$$I_a^{global} = mean(I) \tag{24}$$

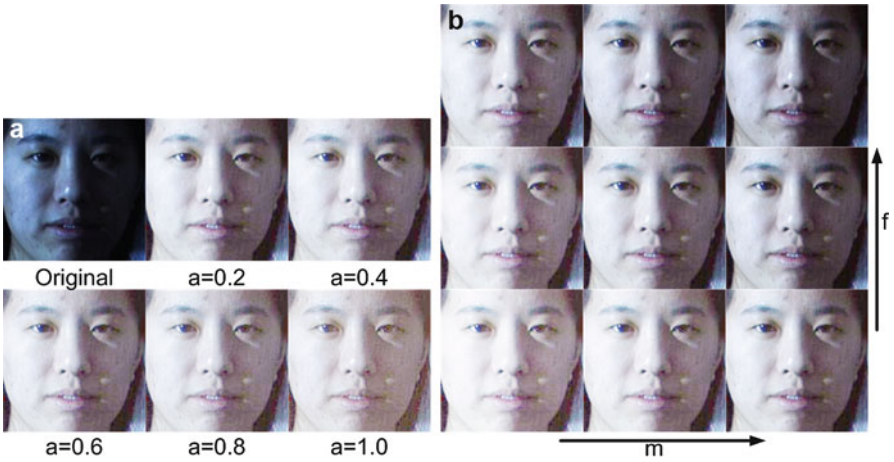

Fig. 6 (**a**) Adaptation level v.s. α parameter, (**b**) lighting adjustment v.s. m and f parameters

Different kernel $K(\bullet)$ can be applied to extract the local information. Gauss kernel is the most commonly used one. The interpolation of the global and local information will adjust the details. In Fig. 6a, with the increasing of the parameter α, the details become notable gradually. When $\alpha = 1$, i.e.$I_a = I_a^{local}$, all the details are expressed out including the noise.

Parameter *f* and *m* The other two parameters f and m control the intensity and contrast, respectively. Parameter f is the multiplier in the adaptation function, i.e. to every point's adaptation level $I_a(x, y)$, f magnifies them with the same scale. The brightness of the whole image will be enhanced or suppressed accordingly.

The alternation of brightness can be shown only when changes on f is large enough. In [41], the parameter f is suggested to be rewritten in the following form

$$f = exp(-f') \tag{25}$$

With a comparative smaller changing range of f', f can alter the brightness of the image.

Parameter m is an exponent in the adaptation function. Different from the parameter f, m magnifies every $I_a(x, y)$ with a different scale based on its adaptation value. Therefore, parameter m can emphasize the difference between every point, i.e. the contrast. In Fig. 6b, the parameter α is fixed. With the increment of m, the contrast of the image is enhanced in every row. And in every column, the brightness of the image is enhanced with the increase of f.

7 Image Sequence Lighting Adjustment

In capturing an image sequence the influence of the scene lighting may not be neglected. Often the variations of the lighting conditions cannot be avoided while recording, and therefore lighting adjustment methods must be used before further processing. In this paper, we propose a tow-steps lighting adjustment approach. First, the initial optimal parameters, α_k^0, f_k^0, m_k^0 of each frame $X_k; k = 1, \cdots N$ are calculated using entropy as an objective function. These values are then used as initial guesses for a constrained least squares optimization problem for further refinement of those parameter. In this step, the objective function is the difference between the adjusted previous frame Y_{k-1} and the current frame X_k. The two steps are detailed in the following sections, and experimental results are presented in Sect. 8.

7.1 Single Image Enhancement

It is well known that an image with large entropy value indicates that the distribution of its intensity values is more uniform, i.e. each intensity value has almost the same probability to appear in the image. Hence, the image cannot be locally too bright or too dark. Entropy $H(x)$, defined as:

$$H(X) = -\sum_{i=0}^{255} p(i) log_2(p(i)) \tag{26}$$

where $p(i)$ is the probability of the intensity values i in the whole image, can be employed to evaluate image lighting quality. When all the intensity values have the same probability in the image, the entropy can reach its maximum value 8. However, not all the images can reach the entropy $H(X) = 8$ when they are in their best situation. The optimal entropy value, H_o, is image content dependent. In this paper, we set $H_o = 7$ as the expected optimal entropy for all the images. Therefore the objective function for the lighting adjustment of every single image is

$$J_1(\alpha, m, f) = \underset{\substack{a\in[0,1];m\in[0.3,1)\\ f\in[\exp(-8),\exp(8)]}}{\arg\min} |H(Y(\alpha, m, f; X)) - H_o| \tag{27}$$

The lighting parameter α controls the adaptation level of the images, as in Eq. (23). It can adjust the image much more than the other two parameters (f, m). Therefore an alternate optimization strategy is used [42]. First, the parameter α is optimized with fixed m and f. Then the parameter m and f are optimized with fixed α. These two optimizations are repeated until convergence. To initialize, we estimate $\hat{\alpha}$ with fixed m and f which are selected according to the luminance situation of

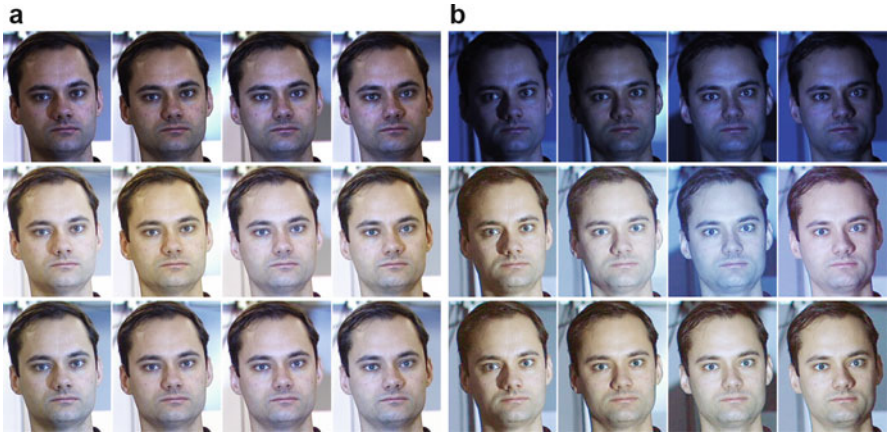

Fig. 7 Lighting adjustment results of frame 1–4 in L1 and L2. (**a**) and (**b**) are results of L1 and L2: *from top to bottom* are original images, entropy-based optimization, and 2-step optimization results, respectively

the image. The contrast-control parameter m can be determined by the key k of the image [41], as

$$m = 0.3 + 0.7k^{1.4} \tag{28}$$

The key of the image evaluates the luminance range of the image and is defined as

$$k = \frac{L_{max} - L_{av}}{L_{max} - L_{min}} \tag{29}$$

where L_{av}, L_{min}, L_{max} are the log average, log minimum and log maximum of the luminance respectively. For color images, we use the luminance image computed as $L = 0.2125I_r + 0.7154I_g + 0.0721I_b$, where I_r, I_g, I_b are the red, green, blue channels. The brightness-control parameter f is set to 1. Then the simplex search algorithm [43] is applied for determining the optimal $\hat{\alpha}$. Fixing the value $\hat{\alpha}$ in J_1, the simplex search algorithm is then used to search for optimal $\hat{m}$ and $\hat{f}$. The alternate optimization will stop when the objective function J_1 is smaller than a given threshold.

This approach can adjust an image to have suitable brightness and contrast. Also, it can enhance the local gradient features of the image due to the adjustment of the parameter α. However, entropy does not relate to intensity directly. Different images can have the same entropy value while their brightness is different. For example, the images in the second row of Fig. 7a, b, being the lighting adjusted results of the images of the first row, have the same entropy values, but their lighting conditions are not similar. Consequently, for a sequence of images, we still need to adjust the brightness and contrast of successive frames to be similar and therefore enhance their features.

7.2 Lighting Adjustment of Successive Images

In video sequences, the difference between successive frames is due to object and/or camera motions and lighting changes. Whereas the former differences are exploited in object tracking and camera motion estimation, the latter, i.e. lighting differences, are such that the required brightness constancy assumption for tracking gets violated. In this paper, we show that for tracking of slow movement in a sequence captured by a fixed camera, the lighting problem can be reduced by applying a lighting adjustment method. Indeed, the lighting of the overall sequence could be made more uniform (in a sequential manner) by considering the changes between successive frames. We propose to use the difference between successive frames as an objective function to estimate the optimal lighting parameters of the current frame X_j, provided that the previous frame X_{j-1} has been adjusted, i.e. given Y_{j-1}:

$$J_2(\alpha, m, f) = \underset{\substack{\alpha\in[0,1];m\in[0.3,1)\\ f\in[\exp(-8),\exp(8)]}}{\arg\min} \sum_x \sum_y \left(Y(\alpha, m, f; X_j(x, y)) - Y_{j-1}\right)^2 \tag{30}$$

With Eq. (20), the difference $e(\alpha, m, f) = Y(\alpha, m, f; X_j) - Y_{j-1}$ between frames can be written as (for simplicity we drop the pixel index (x, y)):

$$e = \frac{X_j}{X_j - (f_j X_{a_j})^{m_j}} - \frac{X_{j-1}}{X_{j-1} - (f_{j-1} X_{a_{j-1}})^{m_{j-1}}} \tag{31}$$

When searching for the optimal parameters for the objective function J_2, the derivatives over different parameters need to calculate.

If the two images concerned are the same, the difference between these images is minimum, at the same time, the difference between the inverse of images will also reach to its minimum value. Therefore, we calculate the difference between the inverse of adjusted images to simplify the computation of derivatives, as

$$\begin{aligned} \tilde{e} &= \frac{X_j - (f_j X_{a_j})^{m_j}}{X_j} - \frac{X_{j-1} - (f_{j-1} X_{a_{j-1}})^{m_{j-1}}}{X_{j-1}} \\ &= \frac{(f_j X_{a_j})^{m_j}}{X_j} - \frac{(f_{j-1} X_{a_{j-1}})^{m_{j-1}}}{X_{j-1}} \end{aligned} \tag{32}$$

Let $\hat{Y}_{j-1} = (f_{j-1} X_{a_{j-1}})^{m_{j-1}} / X_{j-1}$ and apply *log* to both side of Eq. (32), we can simplify the difference between frames further as

$$\begin{aligned} \hat{e} &= \log \frac{(f_j X_{a_j})^{m_j}}{X_j} - \log \hat{Y}_{j-1} \\ &= m_j \log f_j + m_j \log X_{a_j} - \log X_j - \log \hat{Y}_{j-1} \end{aligned} \tag{33}$$

Then the objective function J_2 can be rewritten as

$$\hat{J}_2(\alpha_j, m_j, f_j) = \underset{\substack{\alpha\in[0,1];m\in[0.3,1)\\ f\in[\exp(-8),\exp(8)]}}{\arg\min} \sum_x \sum_y \left(m_j \log f_j + m_j \log X_{a_j} - \log X_j - \log \hat{Y}_{j-1} \right)^2 \quad (34)$$

This formulation allows easily estimating the partial derivatives, and we apply the interior-point algorithm [44] to solve the optimization problem $\hat{J}_2$, with initial values of the lighting parameters α_j^0, f_j^0 and m_j^0 obtained by minimizing Eq. (27).

8 Experiments on Lighting Adjustment

The proposed lighting adjustment algorithms of the previous section have been tested on the PIE facial database [38], from which we selected images under different lighting conditions to compose 3 test sequences, here referred to as L1, L2 and L3. We intend to take these sequences as typical examples to demonstrate the performance of the algorithm in slight lighting variations (L1), overall dark sequences (L2) and suddenly changing light variations (L3). To show the benefits of the proposed image sequence lighting adjustment approach, we compare it to state-of-art lighting adjustment methods for single images, namely, the quotient image (QI) algorithm [4], and the well known histogram equalization (HE) approach.

The lighting conditions of the test sequences can be described as follows. Sequence L1 and L2 are composed of 19 frames taken from the same person. The first row of Fig. 7 shows the first 4 frames of L1 and L2. The images in L1 are taken with ambient lighting and 19 different point light sources. The positions of these light points, are 10, 07, 08, 09, 13, 14, 12, 11, 06, 05, 18, 19, 20, 21, and 22, respectively. The images in L2 are taken under the same light point source but without ambient lighting, so they appear to be more dark. Sequence L3 is composed of 39 images which come from L1 and L2 alternately. Thus the lighting condition of the images in L3 is ambient lighting on and off alternately. The first row of Fig. 9 shows the frames 9–14 of L3.

To evaluate the lighting quality of the adjusted images, the *key value* (Eq. (29)) and entropy are depicted in Fig. 8. The *key value* of an image evaluates the luminance range of the image. The entropy, being the mean entropy of the 3 color channels, relates to the distribution of the intensity values in each channel.

The key value of all adjusted frames and the original sequence of L3 are shown in Fig. 8d. The key value zigzags due to the alternate brightness of the original sequence L3. For a sequence with reduced lighting variation the key value should stay constant throughout the sequence. Therefore, we show the variance of the key value in Fig. 8b. For all the 3 test sequences, the variance of the key value of the results of the proposed 2-step optimization algorithm is smaller than that of the other algorithms except HE algorithm. However, HE algorithm costs the entropy value of images, whose results are even worse than the original images (Fig. 8a). The reason

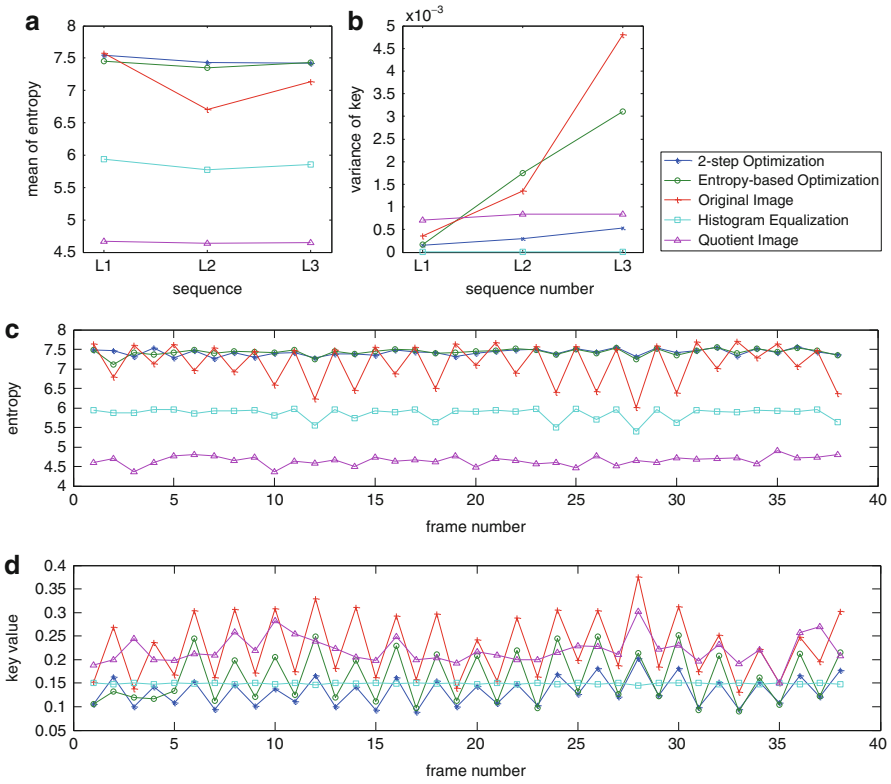

Fig. 8 Entropy and image key curves. (**a**) and (**b**) are the mean entropy and the variance of key of all the frames in the original sequences and adjusted results of the sequences, respectively. (**c**) and (**d**) are the entropy value and key value of every frame in L3 and different adjusted results of L3, respectively

is that HE algorithm can make the intensity distribution uniform only by skipping values in the intensity range [0,255] of the adjusted images, thereby leaving many gaps in the histogram of the adjusted images. The entropy value of the QI results are the smallest because of the loss of the low frequency information in the images. The proposed algorithm is the largest in the mean of entropy, Fig. 8a, and we can also see from Fig. 9a that these resemble most the intensity value distribution of the original images. Our goal is indeed not to change the image appearance dramatically (as compared to QI) but only to obtain a good lighting quality. Therefore, it is quite normal that we couldn't improve L1 sequence so much, which is already captured at a reasonable lighting quality with the ambient light. However, we were still able to adjust its brightness to be more uniform while keeping its high image quality, as shown in Fig. 7a. On the other hand, our 2-step algorithm enhanced the image lighting quality significantly for the sequences L2 and L3 containing images taken under extreme lighting conditions.

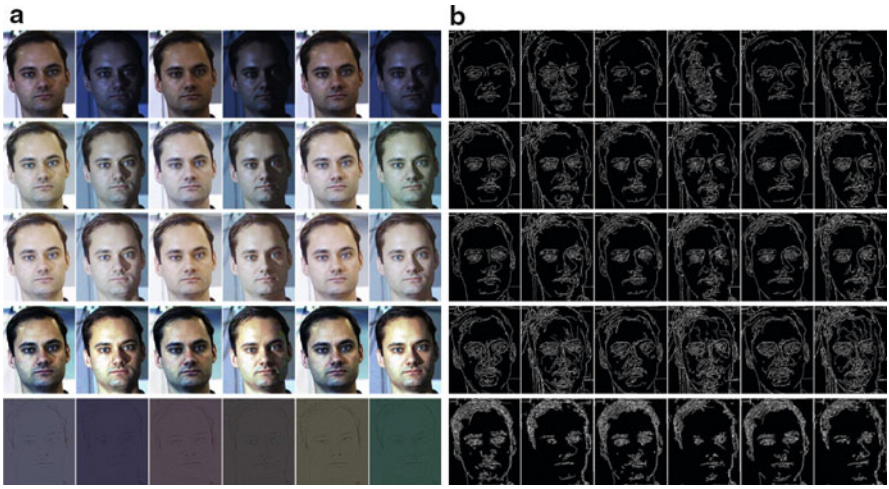

Fig. 9 Lighting adjustment results of frame 9–14 in sequence L3. (**a**) *From top to bottom* are the original images, entropy-based optimization, 2-step optimization, histogram equalization (HE), and quotient image (QI) results, respectively. (**b**) The edge of corresponding images in (**a**)

Next, we examine the effect of the lighting adjustment methods on the object's edges of Fig. 9b to determine if the methods are appropriate as pre-processing for feature detection methods. Considering the edges in the adjusted images, our proposed algorithm enhances the feature of images. This is especially the case for those images taken in a dark environment. Also, highlight are compensated and the influence of shadows on the edges are reduced. The HE algorithm was able to enhance the contrast of the image but at the same time it enhanced noise as well. As we already mentioned, the QI algorithm removed most low frequency information of the image thereby included some important features of the image.

The advantage of the image difference-based optimization step is illustrated for facial feature tracking (on the sequences L1–L3). We demonstrate that the difficulty of tracking a modified object appearance due to lighting changes can be overcome by employing our proposed algorithm as pre-processing. In this paper, we focus on the results of a template-based eye and mouth corner tracker. That tracker is part of a previously developed approach to automatically locate frontal facial feature points under large scene variations (illumination, pose and facial expressions) [45]. This approach consisted of three steps: (1) we use a kernel-based tracker to detect and track the facial region; (2) we constrain a detection and tracking of eye and mouth facial features by the estimated face pose of (1) by introducing the parameterized feature point motion model into a Lukas-Kanade tracker; (3) we detect and track 83 semantic facial points, gathered in a shape model, by constraining the shapes rigid motion and deformation parameters by the estimated face pose of (1) and by the eyes and mouth corner features location of (2).

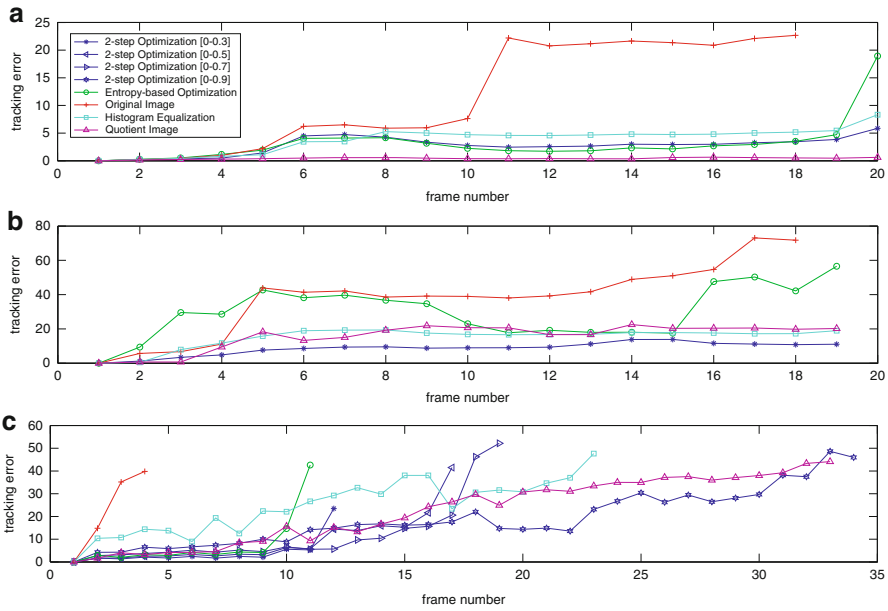

Fig. 10 Feature points tracking error. (**a**)–(**c**) are tracking errors for sequence L1 to L3, respectively

The performance of the tracking of the eyes and mouth corners (6 feature points) on the original and adjusted image sequences L1–L3 is displayed in Fig. 10. The tracking error per frame is calculated as the average distance between the real positions (manually identified) and the tracked positions of the 6 feature points in the image. When tracking was lost, the graph is truncated. Figure 10a shows that all adjustments of the sequence L1 allow to track until the end of that sequence. The QI shows the smallest tracking error because it enhances the gradient features in the image, but at the cost of obtaining visually unpleasant images (see last row of Fig. 9a). Compared to the HE results, our two-step optimization does reach a better tracking performance. Because the initial lighting variations in sequence L1 are not that big, the entropy-step alone may already improve the tracking. The benefit of the image difference-based optimization step becomes obvious via the tracking error graphs of the dark sequence L2 in Fig. 10b. Here, the tracking errors on the 2-step optimization are the smallest. This shows that local features are enhanced very well, but also that taking care of correspondences between images is indeed important. QI and HE adjustments perform worse in tracking. For QI, the reason is that it may enhance the local features (gradients) only when the noise level is not high, i.e. images taken in good lighting conditions such as in L1. On the alternating dark and light sequence L3 the tracking of the original and entropy-optimized sequence is very quickly lost, as shown in Fig. 10c. It is thus crucial to take into account the sequence aspects in lighting adjustment. It is worth noting that the tracking for our proposed algorithm results was lost only when a part of the image were

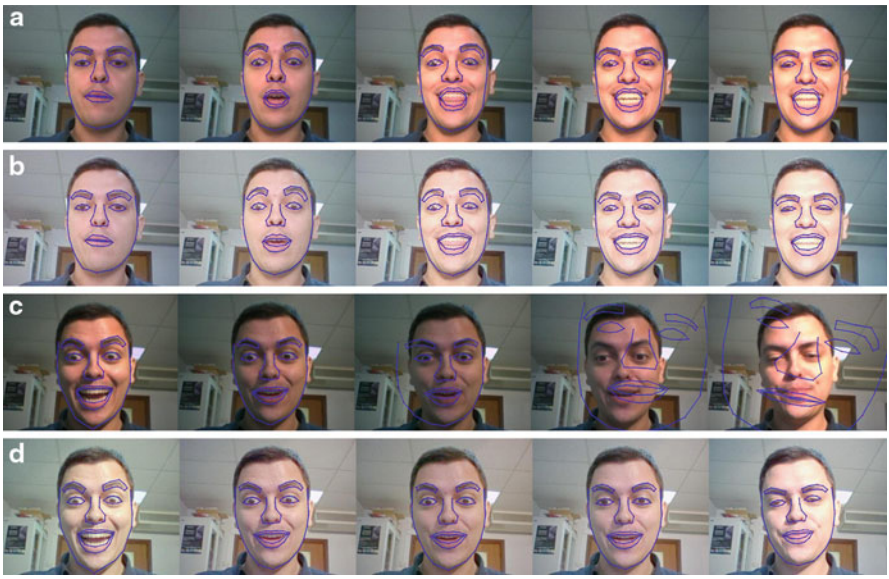

Fig. 11 Shape model results. (**a**) results on original sequence M1 with 10 frames (there are frame 1, 3, 5, 7, 10 *from left to right*), (**b**) corresponding results on adjusted sequence M1; (**c**) results on original sequence M2 with 14 frames (there are frame 1, 4, 6, 9, 14 *from left to right*), (**d**) corresponding results on adjusted sequence M2

in deep shadow (such as frame 12, 17 and 19). Although no adjustment method can track until the end of the sequence, we see that a larger enhancement of the local features may allow to track longer (reduced entropy). That was done by enlarging the alpha range from $[0, 0.3]$ to $[0, 0.9]$ in the 2-step optimization (Eqs. (27) and (30)). When comparing the errors before tracking was lost, we see that reducing frame differences, especially with small alpha range, increases the accuracy of the tracking. This shows that enhancing image sequence quality can also help to track.

Then we tested the constrained shape model tracking (step (3) of [45]) on a sequence [46] adjusted by the 2-step lighting optimization. Before adjustment, shown in Fig. 11a, c, some tracked features could not be well delineated due to the illumination changes in the image sequence. The intensity and texture of the face image were improved by our lighting adjustment and therefore all shape points were tracked more efficiently as shown in Fig. 11b, d.

9 Conclusion

Lighting is always a crucial problem for image based pattern recognition due to the reason that lighting is the main factor that makes the image. Therefore when lighting changes, images of objects will also change such that difficulties arise in detecting, recognizing and tracking them throughout images.

To deal with the practical requirement of few training images, we built a statistical model based on the 9PL theory. With the MAP estimation, we can recover the basis images from one gallery image under arbitrary lighting conditions, which could be single lighting source or multiple lighting sources. The experimental results based on the recovered subspace are comparable to those from other algorithms that require lots of gallery images or the geometric information of the subjects. Even in extreme lighting conditions, the recovered subspace can still appropriately represent lighting variation. The recovered subspace retains the main characteristics of the 9PL subspace. Based on our statistical model, we can build the lighting subspace of a subject from only one gallery image. It avoids the limitation of requiring tedious training or complex training data, such as many gallery images or the geometric information of the subject. After the model has been trained well, the computation for recovering the basis images is quite simple and without the need of 3D models. Besides faces, the proposed algorithm is able to be generalized to model lighting variation for images of other objects in a fixed pose.

Currently, most of the algorithms that deal with the lighting problems are only aimed at adjusting the lighting condition of one image. Furthermore, it is not practical to build lighting models for each object in a complicated scene, especially when the object is unknown beforehand. For the application of image sequences processing, we proposed a 2-step lighting adjustment algorithm to reduce the variations of lighting in an image sequence. First, an entropy-based algorithm is applied to calculate initial lighting parameters of a perceptual lighting model. Then the difference between current and previous frames is employed as an objective function for the further optimization of those lighting parameters. Using this criteria, successive frames are adjusted to have similar brightness and contrast. Image lighting quality, measured by entropy and key value, but also local features are enhanced. The proposed two-step lighting adjustment algorithm can be applied to any image sequences besides facial sequences. We did demonstrate the effectiveness of the proposed algorithm for subsequent image processing, such as detection and tracking.

Acknowledgements This material is based upon work supported by the PhD Programs Foundation of Ministry of Education of China (Grant No. 20136102120041, 20116102120031), National High-tech Research and Development Program of China(863 Program) (No. 2014AA015201), National Natural Science Foundation of China (No. 61103062, No. 61502388), and the Fundamental Research Funds for the Central Universities (No. 3102015BJ(II)ZS016).

References

1. Y. Gao, M.K. Leung, Face recognition using line edge map. IEEE Trans. Pattern Anal. Mach. Intell. **24**(6), 764–779 (2002)
2. H. Zhou, A.H. Sadka, Combining perceptual features with diffusion distance for face recognition. IEEE Trans. Syst. Man Cybern. Part C **41**(5), 577–588 (2011)
3. A. Shashua, T. Riklin-Raviv, The quotient image: class-based re-rendering and recognition with varying illuminations. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 129–139 (2001)

4. H. Wang, S.Z. Li, Y. Wang, Generalized quotient image, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2 (IEEE, New York, 2004), pp. 498–505
5. X. Xie, W.S. Zheng, J. Lai, P.C. Yuen, C.Y. Suen, Normalization of face illumination based on large-and small-scale features. IEEE Trans. Image Process. **20**(7), 1807–1821 (2011)
6. X. Zhao, S.K. Shah, I.A. Kakadiaris, Illumination normalization using self-lighting ratios for 3d2d face recognition, in *European Conference on Computer Vision (ECCV) Workshops and Demonstrations* (Springer, New York, 2012), pp. 220–229
7. Y. Fu, N. Zheng, An appearance-based photorealistic model for multiple facial attributes rendering. IEEE Trans. Circuits Syst. Video Technol. **16**(7), 830–842 (2006)
8. M. De Marsico, M. Nappi, D. Riccio, H. Wechsler, Robust face recognition for uncontrolled pose and illumination changes. IEEE Trans. Syst. Man Cybern. Syst. **43**(1), 149–163 (2013)
9. T. Chen, W. Yin, X.S. Zhou, D. Comaniciu, T.S. Huang, Total variation models for variable lighting face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(9), 1519–1524 (2006)
10. X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Process. **19**(6), 1635–1650 (2010)
11. W. Chen, M.J. Er, S. Wu, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. IEEE Trans. Syst. Man Cybern. B Cybern. **36**(2), 458–466 (2006)
12. X. Xie, K.M. Lam, An efficient illumination normalization method for face recognition. Pattern Recogn. Lett. **27**(6), 609–617 (2006)
13. P.W. Hallinan, A low-dimensional representation of human faces for arbitrary lighting conditions, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, 1994), pp. 995–999
14. S.K. Nayar, H. Murase, Dimensionality of illumination in appearance matching, in *IEEE International Conference on Robotics and Automation*, vol. 2 (IEEE, New York, 1996), pp. 1326–1332
15. P.N. Belhumeur, D.J. Kriegman, What is the set of images of an object under all possible illumination conditions? Int. J. Comput. Vis. **28**(3), 245–260 (1998)
16. A.S. Georghiades, P.N. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 643–660 (2001)
17. S.K. Zhou, G. Aggarwal, R. Chellappa, D.W. Jacobs, Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **29**(2), 230–245 (2007)
18. R. Ramamoorthi, P. Hanrahan, On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object. J. Opt. Soc. Am. A **18**(10), 2448–2459 (2001)
19. R. Basri, D.W. Jacobs, Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Anal. Mach. Intell. **25**(2), 218–233 (2003)
20. C.P. Chen, C.S. Chen, Intrinsic illumination subspace for lighting insensitive face recognition. IEEE Trans. Syst. Man Cybern. B Cybern. **42**(2), 422–433 (2012)
21. A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, Y. Ma, Toward a practical face recognition system: robust alignment and illumination by sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **34**(2), 372–386 (2012)
22. L. Zhuang, A.Y. Yang, Z. Zhou, S.S. Sastry, Y. Ma, Single-sample face recognition with image corruption and misalignment via sparse illumination transfer, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, 2013), pp. 3546–3553
23. L. Li, S. Li, Y. Fu, Discriminative dictionary learning with low-rank regularization for face recognition, in *10th IEEE International Conference Automatic Face and Gesture Recognition* (2013)
24. Y. Zhang, M. Shao, E. Wong, Y. Fu, Random faces guided sparse many-to-one encoder for pose-invariant face recognition, in *International Conference on Computer Vision (ICCV)* (2013)

25. L. Zhang, D. Samaras, Face recognition under variable lighting using harmonic image exemplars, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1 (IEEE, New York, 2003), I–19
26. Z. Wen, Z. Liu, T.S. Huang, Face relighting with radiance environment maps, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2 (IEEE, New York, 2003), II–158
27. L. Zhang, S. Wang, D. Samaras, Face synthesis and recognition from a single image under arbitrary unknown lighting using a spherical harmonic basis morphable model, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2 (IEEE, New York, 2005), pp. 209–216
28. J. Lee, J. Moghaddam, H. Pfister, R. Machiraju, A bilinear illumination model for robust face recognition, in *International Conference on Computer Vision (ICCV)*, vol. 2 (IEEE, New York, 2005), pp. 1177–1184
29. Y. Wang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, D. Samaras, Face re-lighting from a single image under harsh lighting conditions, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, 2007), pp. 1–8
30. X. Zhao, G. Evangelopoulos, D. Chu, S. Shah, I.A. Kakadiaris, Minimizing illumination differences for 3d to 2d face recognition using lighting maps. IEEE Trans. Cybern. **44**(5), 725–736 (2014)
31. K.C. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 684–698 (2005)
32. X. Jiang, Y.O. Kong, J. Huang, R. Zhao, Y. Zhang, Learning from real images to model lighting variations for face images, in *European Conference on Computer Vision (ECCV)* (2008), pp. 284–297
33. X. Jiang, P. Fan, I. Ravyse, H. Sahli, J. Huang, R. Zhao, Y. Zhang, Perception-based lighting adjustment of image sequences, in *Asian Conference on Computer Vision (ACCV)* (2009), pp. 118–129
34. T. Sim, T. Kanade, Combining models and exemplars for face recognition: an illuminating example, in *Proceedings of the CVPR 2001 Workshop on Models versus Exemplars in Computer Vision*, vol. 1 (2001)
35. Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 721–732 (1997)
36. C.G. Atkeson, A.W. Moore, S. Schaal, Locally weighted learning for control, in *Lazy Learning* (Springer, New York, 1997), pp. 75–113
37. L.L. Scharf, *Statistical Signal Processing*, vol. 98 (Addison-Wesley, Reading, 1991)
38. T. Sim, S. Baker, M. Bsat, The cmu pose, illumination, and expression (pie) database, in *Fifth IEEE International Conference on Automatic Face and Gesture Recognition* (IEEE, New York, 2002), pp. 46–51
39. A.M. Martinez, The ar face database. CVC Technical Report, vol. 24 (1998)
40. H. Han, S. Shan, X. Chen, W. Gao, A comparative study on illumination preprocessing in face recognition. Pattern Recogn. **46**(6), 1691–1699 (2013)
41. E. Reinhard, K. Devlin, Dynamic range reduction inspired by photoreceptor physiology. IEEE Trans. Vis. Comput. Graph. **11**(1), 13–24 (2005)
42. X. Jiang, P. Sun, R. Xiao, R. Zhao, Perception based lighting balance for face detection, in *Asia Conference on Computer Vision, ACCV06* (2006), pp. 531–540
43. J.A. Nelder, R. Mead, A simplex method for function minimization. Comput. J. **7**, 308–313 (1965)
44. R.A. Waltz, J.L. Morales, J. Nocedal, D. Orban, An interior algorithm for nonlinear optimization that combines line search and trust region steps. Math. Program. **107**(3), 391–408 (2006)
45. Y. Hou, H. Sahli, I. Ravyse, Y. Zhang, R. Zhao, Robust shape-based head tracking, in *Proceedings of the Advanced Concepts for Intelligent Vision Systems (ACIVS 2007)*, eds. by J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders. Springer Lecture Notes in Computer Science, vol. 4678 (2007), pp. 340–351
46. F. Dornaika, F. Davoine, Simultaneous facial action tracking and expression recognition in the presence of head motion. Int. J. Comput. Vis. **76**(3), 257–281 (2008)